

# **CASE STUDY ABOUT COVID VACCINATION STATE WISE IN INDIA**

## **ABSTRACT**

Free vaccination against COVID-19 commenced in India on January 16, 2021, and the government is urging all of its citizens to be immunized, in what is expected to be the largest vaccination program in the world. Out of the eight COVID-19 vaccines that are currently under various stages of clinical trials in India, four were developed in the country. India's drug regulator has approved restricted emergency use of Covishield (the name employed in India for the Oxford-AstraZeneca vaccine) and Covaxin, the home-grown vaccine produced by Bharat Biotech. Indian manufacturers have stated that they have the capacity to meet the country's future needs for COVID-19 vaccines. The manpower and cold-chain infrastructure established before the pandemic are sufficient for the initial vaccination of 30 million healthcare workers. The Indian government has taken urgent measures to expand the country's vaccine manufacturing capacity and has also developed an efficient digital system to address and monitor all the aspects of vaccine administration.

In the first phase, all healthcare and frontline workers were administered the vaccine. In the second phase, people above the age of 60 and those between 45 and 59 years with comorbidities were vaccinated. From April 1, the government will vaccinate all above 45 years.

## **INTRODUCTION**

India, which has a robust vaccine development program, not only plans for domestic manufacture of COVID-19 vaccine but also for its distribution in countries that cannot afford to buy expensive vaccines from the Western world. In India.

India with its estimated population of 1380 million (as of 2020) is planning to administer the vaccine to all its citizens who are willing to take it. Importation of vaccines might not be the best option for India due to its large population.

Now we are analysing the data which consists of vaccination details of people who are vaccinated in India.

# DATA ANALYSIS

Considering a data set named “covid\_vaccine\_statewise.csv”.

We began by importing the libraries we are going to need. And loading the data set.

## Importing libraries and loading data set

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
In [2]: d=pd.read_csv('covid_vaccine_statewise.csv')
```

```
In [3]: d
```

```
Out[3]:
```

	Updated On	State	Total Individuals Vaccinated	Total Sessions Conducted	Total Sites	First Dose Administered	Second Dose Administered	Male(Individuals Vaccinated)	Female(Individuals Vaccinated)	Total Covaxin Administered	Total CoviShield Administered	Total D Administered
0	16-01-2021	India	48276.0	3455.0	2957.0	48276.0	0.0	23757.0	24517.0	579.0	47697.0	48276.0
1	17-01-2021	India	58604.0	8532.0	4954.0	58604.0	0.0	27348.0	31252.0	635.0	57969.0	58604.0
2	18-01-2021	India	99449.0	13611.0	6583.0	99449.0	0.0	41361.0	58083.0	1299.0	98150.0	99449.0
3	19-01-2021	India	195525.0	17855.0	7951.0	195525.0	0.0	81901.0	113613.0	3017.0	192508.0	195525.0
4	20-01-2021	India	251280.0	25472.0	10504.0	251280.0	0.0	98111.0	153145.0	3946.0	247334.0	251280.0
...	...	...	...	...	...	...	...	...	...	...	...	...
5320	01-06-2021	West Bengal	11075394.0	781523.0	2249.0	11075394.0	3893763.0	6194489.0	4879421.0	1677903.0	13291254.0	14969137.0
5321	02-06-2021	West Bengal	11283299.0	790334.0	1988.0	11283299.0	3903113.0	6317601.0	4964153.0	1693274.0	13493138.0	15186042.0
5322	03-06-2021	West Bengal	11514760.0	803979.0	2104.0	11514760.0	3915654.0	6452599.0	5060564.0	1712330.0	13718084.0	15430914.0
5323	04-06-2021	West Bengal	11795260.0	937852.0	2500.0	11795260.0	3931230.0	6615547.0	5178013.0	1749972.0	13976518.0	15726030.0
5324	05-06-2021	West Bengal	12090072.0	981547.0	2517.0	12090072.0	3941080.0	6784722.0	5303588.0	1806377.0	14224775.0	16031157.0

5325 rows x 12 columns

## d.shape

It shows how many rows and columns are present in the data set.

```
In [5]: d.shape
```

```
Out[5]: (5325, 12)
```

```
In [ ]:
```

```
In [ ]:
```

This data set having 5325 rows and 12 columns.

## d.info()

It returns range, column, number of non-null objects of each column, datatype and memory usage.

```
In [4]: d.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5325 entries, 0 to 5324
Data columns (total 12 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Updated On                            5325 non-null   object
1   State                                 5325 non-null   object
2   Total Individuals Vaccinated          5286 non-null   float64
3   Total Sessions Conducted              5286 non-null   float64
4   Total Sites                           5286 non-null   float64
5   First Dose Administered                5286 non-null   float64
6   Second Dose Administered               5286 non-null   float64
7   Male(Individuals Vaccinated)          5286 non-null   float64
8   Female(Individuals Vaccinated)        5286 non-null   float64
9   Total Covaxin Administered            5286 non-null   float64
10  Total CoviShield Administered          5286 non-null   float64
11  Total Doses Administered               5288 non-null   float64
dtypes: float64(10), object(2)
memory usage: 499.3+ KB
```

## df.describe()

Generate descriptive statistics that summarize the central tendency, dispersion, and shape of a dataset's distribution, excluding NaN values.

```
In [6]: d.describe()
```

Out[6]:

	Total Individuals Vaccinated	Total Sessions Conducted	Total Sites	First Dose Administered	Second Dose Administered	Male(Individuals Vaccinated)	Female(Individuals Vaccinated)	Total Covaxin Administered	Total CoviShield Administered	Total Doses Administered
count	5.286000e+03	5.286000e+03	5286.000000	5.286000e+03	5.286000e+03	5.286000e+03	5.286000e+03	5.286000e+03	5.286000e+03	5.288000e+03
mean	3.701878e+06	2.762206e+05	2370.913924	3.688361e+06	8.140408e+05	1.950188e+06	1.751187e+06	4.445988e+05	4.041478e+06	4.500656e+06
std	1.506484e+07	1.083285e+06	7583.357655	1.505896e+07	3.620258e+06	7.964782e+06	7.100591e+06	1.921085e+06	1.670661e+07	1.861008e+07
min	7.000000e+00	0.000000e+00	0.000000	7.000000e+00	0.000000e+00	0.000000e+00	2.000000e+00	0.000000e+00	7.000000e+00	0.000000e+00
25%	6.053525e+04	2.978250e+03	67.000000	6.029875e+04	2.136500e+03	3.161100e+04	2.704500e+04	0.000000e+00	6.847800e+04	6.861175e+04
50%	3.230380e+05	1.524800e+04	591.500000	3.139450e+05	5.604050e+04	1.597780e+05	1.604310e+05	1.251500e+03	3.366755e+05	3.686250e+05
75%	2.723301e+06	1.886700e+05	1825.750000	2.723301e+06	5.069208e+05	1.432426e+06	1.273246e+06	2.701952e+05	2.868969e+06	3.209040e+06
max	1.869338e+08	1.313116e+07	73933.000000	1.869338e+08	4.512972e+07	1.006564e+08	8.624609e+07	2.703112e+07	2.050146e+08	2.320635e+08

## df.info()

It returns range, column, number of non-null objects of each column, datatype and memory usage.

```
In [6]: d.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5325 entries, 0 to 5324
Data columns (total 12 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   Updated On                           5325 non-null   object
1   State                                5325 non-null   object
2   Total Individuals Vaccinated          5286 non-null   float64
3   Total Sessions Conducted             5286 non-null   float64
4   Total Sites                          5286 non-null   float64
5   First Dose Administered               5286 non-null   float64
6   Second Dose Administered              5286 non-null   float64
7   Male(Individuals Vaccinated)          5286 non-null   float64
8   Female(Individuals Vaccinated)        5286 non-null   float64
9   Total Covaxin Administered            5286 non-null   float64
10  Total CoviShield Administered          5286 non-null   float64
11  Total Doses Administered              5288 non-null   float64
dtypes: float64(10), object(2)
memory usage: 499.3+ KB
```

## df.isnull().sum()

It returns a number of null values in each column.

```
In [5]: d.isnull().sum()

Out[5]: Updated On                0
        State                    0
        Total Individuals Vaccinated    39
        Total Sessions Conducted       39
        Total Sites                  39
        First Dose Administered        39
        Second Dose Administered       39
        Male(Individuals Vaccinated)   39
        Female(Individuals Vaccinated) 39
        Total Covaxin Administered     39
        Total CoviShield Administered  39
        Total Doses Administered       37
        dtype: int64
```

Each column consisting null values except “updates on and state” column.

So, we need to remove null from those columns to avoid inaccuracy in the prediction. We use the Imputer from sklearn.impute import SimpleImputer. To fill the missing values in every column with the mean.

```
In [11]: c=['Total Individuals Vaccinated',
            'Total Sessions Conducted',
            'Total Sites ',
            'First Dose Administered',
            'Second Dose Administered',
            'Male(Individuals Vaccinated)',
            'Female(Individuals Vaccinated)',
            'Total Covaxin Administered', 'Total CoviShield Administered', 'Total Doses Administered']
```

```
In [12]: from sklearn.impute import SimpleImputer
          # invoking SimpleImputer to fill missing values
          imputer = SimpleImputer(missing_values=np.nan, strategy='mean')
          d[c] = imputer.fit_transform(d[c])
```

Now again we need to check whether the null values are there or not in the data set after the completion of above operation.

```
In [13]: d.isnull().sum()
```

```
Out[13]: Updated On          0
          State              0
          Total Individuals Vaccinated  0
          Total Sessions Conducted    0
          Total Sites          0
          First Dose Administered    0
          Second Dose Administered   0
          Male(Individuals Vaccinated)  0
          Female(Individuals Vaccinated) 0
          Total Covaxin Administered   0
          Total CoviShield Administered 0
          Total Doses Administered     0
          dtype: int64
```

The data set consists of no null values in it.

Now we can see some of the visualizations. By using the library matplotlib and seaborn. There are mainly three types. They are:-

Line chart

Scatter plot

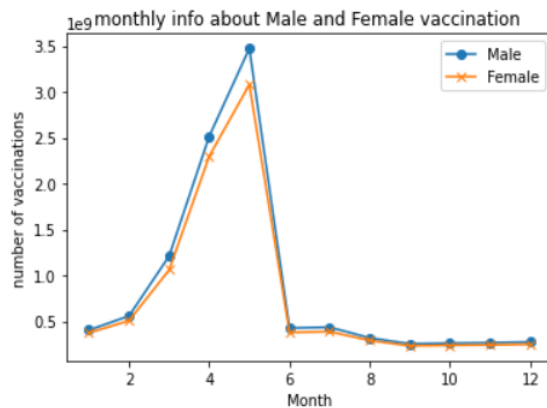
Bar chart

## Line chart

```
In [98]: monthly_male = d.groupby('month')[['male']].sum()
```

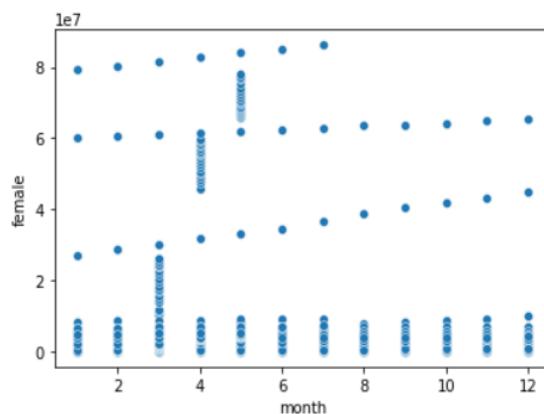
```
In [99]: monthly_female = d.groupby('month')[['female']].sum()
```

```
In [100]: plt.plot(monthly_male,marker='o')
plt.plot(monthly_female,marker='x')
plt.xlabel('Month')
plt.ylabel('number of vaccinations')
plt.legend(['Male', 'Female'])
plt.title("monthly info about Male and Female vaccination ");
```



## Scatter plot

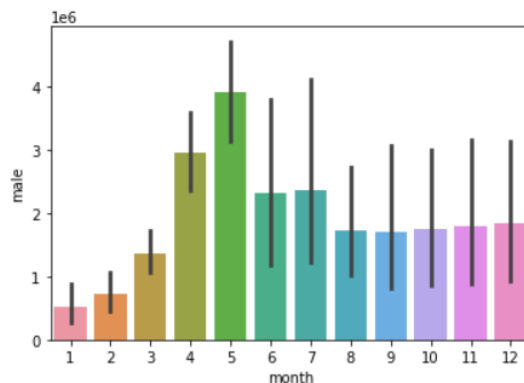
```
In [110]: sns.scatterplot(x=dframecopy.month, y=dframecopy.female);
```



## Bar chart

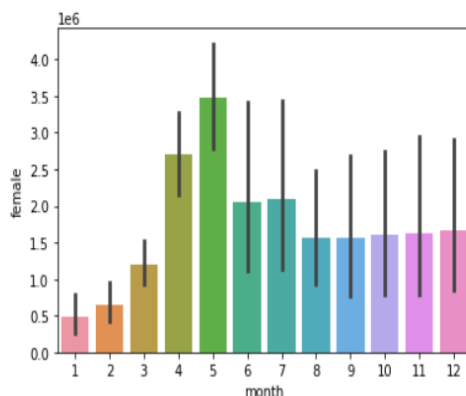
It shows the information about male vaccinations month wise.

```
In [91]: sns.barplot(x='month',y='male',data=dframecopy);
```



Now the female vaccination graph month wise.

```
In [92]: sns.barplot(x='month',y='female',data=dframecopy);
```



## CONCLUSION

In this data set we have seen the analysis of the given data.

COVID-19, a new and sometimes deadly respiratory illness that is believed to have originated in a live animal market in China, has spread rapidly throughout that country and the world. As per the suggestions of doctors by getting

vaccination done we can control the covid-19. This leads to the obtaining of less number of cases.

## REFERENCES

Dataset collected from : [https://www.kaggle.com/sudalairajkumar/covid19-in-india?select=covid\\_vaccine\\_statewise.csv](https://www.kaggle.com/sudalairajkumar/covid19-in-india?select=covid_vaccine_statewise.csv)

Learned this course from : <https://jovian.ai/learn/data-analysis-with-python-zero-to-pandas>

by

LIKHITHA AARAGA